Numerical Weather Prediction

Forecasting Research Scientific Paper No. 69

Modelling of error covariances by the ensemble Kalman filter

By A.C. Lorenc

June 2002

© Crown Copyright 2002

Met Office , NWP Division , Room 344, London Road , Bracknell , Berkshire ,RG12 2SZ, United Kingdom

Numerical Weather Prediction

NWP

Forecasting Research Scientific Paper No. 69

Modelling of error covariances by the ensemble Kalman filter

June 2002

Andrew C. Lorenc

© Crown Copyright 2002

(Submitted to the *Quarterly Journal of the Royal Meteorological Society*)

Met Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 2002

Permission to quote from this paper should be obtained from the Director NWP at the above address.

For additional copies, or information about other NWP publications, contact
john.armstrong@metoffice.com
Tel: 44 (0)1344 856245
Fax: 44(0)1344 854026

Modelling of error covariances by the ensemble Kalman filter

Andrew C. Lorenc

Summary

The ensemble Kalman (EnKF) filter method is summarised and reviewed for its applicability to data assimilation for numerical weather prediction (NWP). Algorithms with sequential processing of observations, using either perturbed observations, or an ensemble adjustment Kalman filter, should be relatively easy to implement, and affordable in computer time, compared to the currently more popular four-dimensional variational assimilation (4D-Var). The use of an existing forecast model, the analysis algorithm, and parallel computing, are all simpler than 4D-Var, but limited-area versions are more complicated.

The EnKF represents the probability distribution of forecast errors using a covariance from a relatively small sample. The noise this introduces is potentially serious, but can be ameliorated by localising the covariances using a Schur product. The ensemble size also limits the number of observed pieces of information which can be fitted in a region. The Schur product removes the ability to impose balance through the covariances, so a separate initialisation may be needed. Good variances should improve the quality control of observations against the forecast, and techniques to do this can be incorporated into the sequential EnKF algorithm. But this is sub-optimal with regard to checking an observation against its neighbours. Similarly the EnKF is sub-optimal in its use of observations nonlinearly related to model variables. For both these aspects variational methods work well.

As a compromise, a variational algorithm is proposed using a hybrid of the standard variational covariance model, and the EnKF covariances including the Schur product. This should enable the good attributes of both methods to be retained, at the expense of having a system more complex than either.

For the reasons discussed, the EnKF may be attractive when building a new medium-range ensemble NWP system. However it remains to be demonstrated that it can use high-resolution satellite data as effectively as 4D-Var. For limited-area mesoscale NWP systems the hybrid method is more attractive.

Keywords: ensemble Kalman filter, numerical weather prediction, error covariance modelling

1. Introduction

Numerical weather prediction (NWP) users are becoming more sophisticated in their ability to allow for the probable range of forecast errors when making decisions. Increased computing power has enabled the use of ensembles as a practical way of representing the probability distribution of errors, and of propagating that distribution in time. This paper discusses whether the ensemble representation of a pdf can be sufficiently accurate as to enable near-optimal combination of forecast and observed information in an ensemble based data assimilation scheme (e.g. Houtekamer and Mitchell 2001, Bishop *et al.* 2001, Anderson 2001), and compare the EnKF approach to the currently more popular variational approach. Variational data assimilation has been a

very successful technique in operational numerical weather prediction (NWP), first in a three-dimensional form (3D-Var: e.g. Parrish and Derber 1992, Courtier *et al.* 1998, Lorenc *et al.* 2000), then using an adjoint model to include the time dimension (4D-Var: e.g. Rabier *et al.* 2000). The key to its success is the emphasis on characterisation of the error structures (and hence information content) of both the model forecast and the observations, enabling them to be combined in a nearly optimal way. This is important since model forecasts used in data assimilation are now typically as accurate as radiosonde observations (Simmons and Hollingsworth 2002), while satellite soundings have error characteristics which can be more easily represented in the direct variational assimilation of radiances (e.g. Andersson *et al.* 1994). The complete error structure of a time-evolving model trajectory requires a multi-dimensional probability distribution function (pdf) which is impossible to know or even represent accurately. Lorenc (2002) discusses how 4D-Var summarises it by a covariance model including time-evolution within a short time-window by a linear perturbation model and its adjoint. The ensemble approach on the other hand represents the pdf less precisely, but can calculate its evolution over a longer time-period.

Lorenc (2002) related 4D-Var to the Kalman Filter. In section 2 I review the basic Ensemble Kalman Filter (EnKF) method in a similar notation. The EnKF is in principle simple to implement; this is discussed in section 3. But ease of implementation has to be weighed against effectiveness; in section 4 I discuss the effect of the EnKF covariances on the characteristics of the assimilation scheme. Although the standard EnKF, reviewed in section 2, is implemented as an observation-space algorithm, this is not essential. In section 5 I describe a hybrid scheme, using the ensemble covariance model in a standard model-space variational algorithm. This is less convenient for generating the ensemble, but might be a useful compromise if a forecast ensemble is generated by other methods. Finally in section 6 I summarise sections 3 and 4 with the help of Table 1 and Table 2, comparing the EnKF with variational methods and discussing the implications for developments to operation NWP assimilation systems.

2. *Review of Ensemble Kalman Filter method*

(a) Equations

The basic idea of the EnKF (Evensen 1994) is to construct an ensemble of forecast states $\{\mathbf{x}_i^f\}(i=1,\ldots,N)$[1] such that the mean of the ensemble is the best estimate of the population mean, and the sample covariance is a good estimate of the forecast error covariance of this best estimate:

$$\mathbf{P}^f \approx \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T}, \tag{1}$$

where the overbar denotes the ensemble estimate of the population mean. As discussed in Lorenc (2002), we follow the extended Kalman filter in representing the forecast error

[1] I follow the normal notation and use *i* as a subscript denoting the ensemble member. This should not be confused with its use to denote the time in the notation of Ide *et al.* (1997) and Lorenc (2002), for which in this paper I use *k*.

pdf by its mean and covariance, even though the forecast states are propagated in time by a nonlinear forecast model M. A similar approximation involving the nonlinear observation operator H allows us to calculate estimates of covariances used in the Kalman gain without needing its linearisation $\mathbf{H}$:

$$\mathbf{P}^f\mathbf{H}^T \approx \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(H(\mathbf{x}^f) - \overline{H(\mathbf{x}^f)}\right)^T},$$

(2)

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T \approx \overline{\left(H(\mathbf{x}^f) - \overline{H(\mathbf{x}^f)}\right)\left(H(\mathbf{x}^f) - \overline{H(\mathbf{x}^f)}\right)^T}.$$

(3)

These can be used in the standard Kalman analysis equation to update the best estimate (ensemble mean):

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\overline{\mathbf{x}^f}\right)\right).$$

(4)

The simplest way to generate the ensemble is to introduce stochastic perturbations into the forecast and analysis of each member:

$$\mathbf{x}_i^f(t_{k+1}) = M\left(\mathbf{x}_i^a(t_k)\right) + \eta_i(t_k),$$

(5)

where $\boldsymbol{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}$, and

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\mathbf{x}_i^f\right) + \varepsilon_i\right),$$

(6)

where $\boldsymbol{\varepsilon}$ is a noise process with zero mean and covariance matrix $\mathbf{R}$. For a sufficiently large ensemble this EnKF gives the same error statistics as the standard Kalman filter (Burgers *et al.* 1998). Note that although (6) must be solved for each ensemble member, the term $\mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}$ is common, and need only be evaluated once.

It is convenient to introduce a matrix notation for the ensemble, with $\mathbf{X}$ having as columns the normalised deviations from the ensemble mean:

$$\mathbf{X} = \frac{1}{\sqrt{N-1}}\left(\mathbf{x}_1 - \overline{\mathbf{x}} \quad \mathbf{x}_2 - \overline{\mathbf{x}} \quad \cdots \quad \mathbf{x}_N - \overline{\mathbf{x}}\right).$$

(7)

These perturbation matrices are estimates of the square-root of the covariances, for instance **(1)** can be replaced by

$$\mathbf{P}_e^f = \mathbf{X}^f \mathbf{X}^{fT}.$$

(8)

The EnKF is thus a square-root Kalman filter; the analysis method (6) allow us to calculate $\mathbf{X}^a$ from $\mathbf{X}^f$, and the forecast (5) allows us to propagate the square-root of the covariance to the next time.

Instead of adding the noise in (6), it is possible to transform the ensemble $\{\mathbf{x}_i^f\}$ with covariance $\mathbf{P}^f$ into an ensemble $\{\mathbf{x}_i^a\}$ with the expected Kalman filter analysis error covariance $\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}^f$. First the ensemble mean is calculated using (4), and then the perturbation matrices are scaled to give the correct covariance. In the Ensemble Adjustment Kalman Filter (EAKF, Anderson 2001) an adjustment matrix $\mathbf{A}$ is implicitly calculated[2] by using a singular value decomposition to find the square-root of a matrix of order the number of observations, such that

$$\mathbf{X}^a = \mathbf{A}^T\mathbf{X}^f.$$

(9)

The Ensemble Transform Kalman Filter (ETKF, Bishop *et al.* 2001) requires the singular value decomposition of a matrix on order N, to find matrix $\mathbf{T}$ such that

$$\mathbf{X}^a = \mathbf{X}^f\mathbf{T}.$$

(10)

(b) Sequential application

The analysis step (6) (or the alternative (4) and (9)) can be applied sequentially to small batches of observations (Houtekamer and Mitchell 2001). The analysed ensemble $\{\mathbf{x}_i^a\}$ from one batch becomes the $\{\mathbf{x}_i^f\}$ for the next. As long as there is no error correlation between observations in different batches, this is equivalent to treating the observations in one large batch. This technique, together with the modification of the covariances to give compact support discussed in 4(a), makes the basic EnKF and the EAKF computationally feasible for large systems.

(c) Application to a time-window

With appropriate changes to H to include an interpolation in time, all of the above equations still hold if we replace each three-dimensional forecast state $\mathbf{x}$ by a four-dimensional $\underline{\mathbf{x}}$ such that $\underline{\mathbf{x}}^T = \left(\mathbf{x}(t_1)^T \quad \cdots \quad \mathbf{x}(t_k)^T\right)$. The covariances given by equations like (1) are then defined between all points in space and time spanned by $\underline{\mathbf{x}}$. This could

[2] In practice the multiplication in (9) is performed without explicitly evaluating $\mathbf{A}$.

allow the batching of observations into time-windows, as in standard NWP assimilation cycles, without the error of assuming that all observations in a time-window are simultaneous. (Although there is little advantage in this for the sequential schemes described in (b) above, and the costs of updating the ensembles increase with the size of $\underline{\mathbf{x}}$.) The time covariances are key to the use of the ETKF method for targeting adapting observations (Bishop *et al.* 2001); they are used to calculate the decrease in error at the end of the time period due to various distributions of observations at the beginning.

(d) Inbreeding

For a small ensemble size, sampling errors in the covariance estimate can build up to give an ensemble whose covariance is too small. To counter this, Anderson (2001) adds an extra covariance inflation factor into (9), while Houtekamer and Mitchell (2001) split the ensemble into two, such that the analysis of one half is performed using the covariances estimated by the other. Other effects of a small ensemble size are discussed in section 4.

3. Ease of implementation of EnKF and 4D-Var

(a) Forecast model

It is possible to use a forecast model in the EnKF which represents all processes thought to be important. Each ensemble state represents a possible atmospheric state, so conceptually simple physical ideas can be used in the design of each parametrisation, and of the stochastic forcing terms representing uncertainties in the model formulation. No linearisation, or adjoint model, is needed.

In comparison the full model used in 4D-Var should evolve the ensemble mean, which may not be a plausible physical state, and parametrisations should allow for the effect of perturbations about this state. Fast processes (compared to the time-window of the 4D-Var), whose timing is uncertain, are particularly hard to treat. In practice smoothly evolving forecast models do not exist, and approximations have to be made in the linearisation and adjoint. Processes which have proved difficult include boundary layer fluxes, and evaporation of falling precipitation - both of which can proceed rapidly then saturate. There is no problem in including a straightforward physical parametrisation of these processes in the EnKF.

(b) Analysis algorithm

A key part of a variational data assimilation system is the background error covariance model, considerable effort is expended designing and coding this. The EnKF is simpler to design and code; the major problem being to keep the ensemble covariance matching the error. To achieve this the twin ensemble approach of Houtekamer and Mitchell (2001) may be used, and some testing and tuning of stochastic parametrisations in the model, and covariance inflation factor is needed.

In comparison to that for the forecast model, little effort needed to produce linear and adjoint versions of the observation operators for variational schemes. The effort is probably worth it for the extra flexibility variational schemes have in dealing with nonlinear observations (discussed in 4(f) below); the EnKF is inferior in this regard. An exception to this might be precipitation; it would be attractive to be able to assimilate

precipitation observations in an EnKF without needing to worry about linear adjoints. (But an EnKF using precipitation observations has yet to be demonstrated.)

(c) Parallel computing

If 4D-Var uses O(100) iterations, running its linear and adjoint model, and the EnKF uses O(100) ensemble members, their computational costs should be similar. The 4D-Var runs have to be performed sequentially, so the linear and adjoint models have to be designed to parallelise on multi-processor computers. In practice this limits the number of processors that can be effectively used. A similar strategy can be adopted for the model in the EnKF, so it cannot be less easy to parallelise. Alternatively the forecast model runs can be distributed over processors, making the forecast step easily scalable to large ensemble sizes and numbers of processors (Keppenne 2000). However a costly transposition of the ensemble matrix is then needed in order to distribute the analysis of batches of observations over the processors. If frequent observations are to be assimilated, these transposition costs may dominate.

(d) Limited-area modelling

In order to provide high resolution over an area of interest, most NWP centres run limited-area forecast models, with lateral boundary conditions specified from a global model. Covariance models in variational schemes can be extended to specify plausible errors in these. Despite theoretical shortcomings in the specification of boundary conditions, in practice such NWP systems give useful short-range forecasts, away from the boundary. EnKF methods however depend on the ensemble members evolving realistic error structures over an extended assimilation cycle. During this time information can propagate from the boundary across the entire area. For a small area, forecast errors will be more influenced by boundary errors than by analysis errors and error growth within the area. So proper specification of an ensemble of boundary values, characterising the incoming errors, would be essential for a limited-area EnKF. Probably this requires an ensemble of global models providing the boundary conditions.
Although research is underway into the use of limited-area ensembles for forecasting, it has concentrated on perturbing the boundary values, and the model formulations, with no published work on the EnKF.

4. Assimilation characteristics of EnKF

(a) The effect of noise in sampled covariances

If the ensemble is created by random sampling from a pdf, covariances estimated using (1) will have error proportional to $1/N$, as illustrated in figure 1, which shows sampled covariances from a Gaussian shaped covariance $\exp\left(-r^2/2a^2\right)$, with $a = 500$ km. A sub-optimal Kalman gain calculated using the estimated covariances:

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{11}$$

will give larger analysis errors, which can be calculated if one knows the true covariance:

$$\mathbf{P}^a = \mathbf{P}^f - 2\mathbf{K}_e\mathbf{H}\mathbf{P}^f + \mathbf{K}_e\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}_e^T. \qquad (12)$$

Applying this to a global analysis using a single perfect observation and the covariances shown in figure 1, gives an average analysis error of 1+0.0087, larger than the prior variance of 1.0. Perfect covariances (e.g. an infinite ensemble size in our example) reduce the global variance to 1-0.0015. In other words the harm done by spurious covariances with distant grid-points is greater than the local benefit. One remedy is to do many approximate local analyses, only selecting local observations. This can introduce discontinuities in the analysis when an observation is selected for one grid-point, but not for its neighbour. Houtekamer and Mitchell (2001) avoid discontinuities by modifying the covariances to make them "local", using a Schur product, with a correlation function with compact support. (The Schur product of two matrices having the same dimensions is denoted $\mathbf{A} = \mathbf{B} \circ \mathbf{C}$ and consist of the element-wise product such that $A_{i,j} = B_{i,j}C_{i,j}$. If $\mathbf{B}$ and $\mathbf{C}$ are valid covariance matrices, then so is $\mathbf{A}$.) I follow them and use as $\mathbf{C}$ the correlation function (4.10) from Gaspari and Cohn (1999), which goes to zero at distance $2c$. This is applied as a simple function of horizontal distance, to all levels and variables. Gaspari and Cohn (1999, figure 6) show that (4.10) is very similar to the Gaussian shaped covariance assumed in figure 1. By trying different values of c, and calculating the analysis error for each using (12), an optimal value can be found. The results are shown in figure 2, for different ensemble sizes. The scales are shown relative to that which matches the curve in figure 1: $c = a(0.3)^{-½} = 913km$, since then results are not sensitive to the value chosen for a. The covariances modified with the Schur product using the optimal c are shown in figure 3.

The use in $\mathbf{C}$ of a compact support function which is zero beyond a known distance allows significant computational savings, for instance in the $\mathbf{P}^f\mathbf{H}^T$ term of (6) and the $\mathbf{A}^T$ term of (9). (The ETKF method transforms the matrix problem to that of determining $\mathbf{T}$ in (10) of order the ensemble size N, rather than working in the space where $\mathbf{C}$ is defined. Because of this, it is not possible to include a Schur product in the ETKF method as currently defined).

(b) The effect of degenerate covariances

The basic EnKF (without the Schur product) only has N degrees of freedom available to fit the observations. This is immediately obvious from the variational formulation of section 5, and from (4) or (10) which have matrices of order N. It is perhaps less apparent from the sequential algorithm (although this can be shown to be equivalent to a simultaneous analysis); appendix A shows that a perfect observation removes a degree of freedom from the ensemble. Thus the covariance given by (1) has at most N non-zero eigenvalues, reduced by one for each recent accurate observation. Although positive definite observational errors $\mathbf{R}$ in theory keep equations well behaved, in practice unrealistic extrapolations into data-sparse areas can result. The EAKF and ETKF algorithms perform singular-value decompositions, and are well behaved if vectors with small singular values are omitted (Anderson 2001). It is clear from section 5 that the

Schur product adds extra degrees of freedom. Houketamer and Mitchell (2001) say that the Schur product greatly improves the conditioning, implying that it was sufficient to keep the system well-behaved for their tests.

A consequence of this degeneracy is that the ensemble Kalman filter is unable to fit more than N data within any area where the effect of the Schur product is insignificant (i.e. where $\mathbf{C}$ is close to 1). Given the limitations on the scale of the Schur product implied by figure 2, this is a serious limitation for some applications. Of particular concern is the ability to fit observations determining the moisture field. Moisture fields often have more horizontal detail than the dynamical fields upon which EnKF experimentation has so far concentrated, and there are observations available (from satellites, and radar) which give frequent detailed information about them. As correlations between temperature and moisture are important they should be analysed together. The horizontal scale of the Schur product must be larger than the scale of the dynamical errors that can affect temperatures even in a mesoscale model, making it difficult to fit many details observed in the moisture field on a smaller scale. (This problem might be alleviated by having a more complicated form for $\mathbf{C}$ than the simple function of horizontal distance used so far. But it may be difficult to specify such a $\mathbf{C}$; the theoretical basis of the EnKF is compromised when $\mathbf{C}$ has a significant affect of the results, rather than mainly reducing noise and computer costs).

(c) Balance

The atmospheric state is attracted towards a slow manifold, with flow close to balance which can be approximated by the geostrophic relationship. As long as the noise introduced in each analysis step is less than the dissipation of imbalance during the forecast, the states $\{\mathbf{x}_i^f\}$ will be close to balance, and this will be reflected by the covariances in (1). In assimilation schemes with local data selection (e.g. Lorenc 1981), changes in data selection from point to point introduce more imbalance than is usually dissipated during a 6 hour forecast to the next batch of observations. It has therefore been usual to include an initialisation step in the data assimilation cycle. The Schur product reduces this data selection noise (Houtekamer and Mitchell 2001). But it still introduces imbalance. Let us assume that the ensemble $\{\mathbf{x}_i^f\}$ is in geostrophic balance, with geopotential height error covariance as in figure 1. Then, without the Schur product, the height increments from a single height observation will approximate the curve in figure 1, and the wind increments will be in geostrophic balance with its gradient. With the Schur product, the height increments will drop of more rapidly with distance, as shown in figure 3, requiring an increased wind increment to balance them, while the actual wind increments implied by the ensemble covariance are reduced by the Schur product factor. Figure 4 quantifies this effect; the wind increments implied by a height observation are significantly sub-geostrophic. It is thus likely that the EnKF will either need an explicit initialisation step, or it will need to use a forecast model with extra damping of gravity waves. This makes it harder to get reliable short-period forecasts of precipitation.

In variational assimilation methods, balance is usually introduced additional "slow" penalties (Gauthier and Thépaut, 2001), or explicitly when designing the covariance model. Lorenc (2002) discussed how the balance constraints are imposed via a null space of the covariance model. As discussed in the previous section, an EnKF covariance has a

spurious null space due to the limited ensemble size. It is impossible to relax this, by the Schur product or otherwise, without relaxing the balance constraints at the same time. Additional initialisation is therefore necessary.

(d) Non-Gaussian errors and nonlinearities

The equations derived in section 2(a) are only optimal if the error pdfs are Gaussian[3]. Nonlinearities in M and H, and the possibility of gross observational errors, can all give non-Gaussian pdfs. Since we have an ensemble to represent the forecast pdf, it is possible to derive alternative methods which do not assume it to be Gaussian. Similarly since we know the properties of H, and of the observational errors, it is possible to allow for a non-Gaussian observational error pdf. However, although possible in principle, in practice there are serious difficulties. Some assumption must be made about the forecast error pdf sampled by the ensemble; none is obviously best. For the sequential processing to be optimal, the ensemble has to represent accurately the key features of the intermediate pdf. For instance if the observation being processed is close to being rejected as erroneous, then the pdf will become bi-modal, and the ensemble should have members distributed in each peak according to its probability. The problems of sampling error seen in 4(a) above will be magnified. Then if a subsequent observation is processed which supports the first, one mode becomes much more likely. If we are to retain the assumption that the ensemble members are equally likely, then some must be moved from one peak to another. This is not easy to achieve; the linear mapping of (9) or (10) has to be replaced by a nonlinear resampling. Such methods have been demonstrated successfully for simple low-order models (J. Anderson, personal communication). But for the large systems used in NWP, it is not currently practicable to replace the main methods of 2(a). Some allowances might still be made for non-Gaussian errors in pre-processing, as discussed in the next two sub-sections.

(e) Quality control

The properties of observations which may have gross errors of can be represented by non-Gaussian distributions with long tails (Lorenc and Hammon 1988, hereafter L&H). Variational methods can cope with these directly (Dharssi *et al.* 1992, Andersson and Jarvinen 1999). These methods allow an implicit choice to be made on whether to believe an outlying observed value, dependent on whether it is consistent with surrounding values. In the algorithms of section 2(a) this is not possible, since the observational error pdfs are only summarised by their covariances.

It is still possible to use the ensemble information in a prior quality control step. Many critical quality control decisions are made in data-sparse areas, where the principal source of corroborative information is the forecast. If the forecast error variance modelled by $\{\mathbf{x}_i^f\}$, despite sampling noise, is more accurate than the error variances assumed in variational schemes, then the ensemble covariances can do better. Figure 5 illustrates this using values typical of ship pressure observations take from L&H. A million observations were generated; 97% of them with Gaussian errors with variance 1 hPa^2,

[3] The equations can also be derived as the best linear unbiased estimator (BLUE) without making assumptions about the form of the pdfs, but for non-Gaussian pdfs the optimal equations should be nonlinear.

and 3% with "gross errors", in which case they can have any plausible value with a constant pdf of 0.043 hPa^{-1}. Note that we are assuming prior removal of extreme observations; the bad observations only have errors in the range ±11.6 hPa. So a significant number of the gross errors are in the same range as good observations, and cannot be easily detected. L&H assumed forecast errors were Gaussian, with variance 3.6 hPa^2. It is more realistic to allow for day-to-day changes in the variance, so forecast values were calculated for each observation, sampled from a range of Gaussians with variances spread evenly from 1.8 to 5.4 hPa^2. Figure 5(a) shows the result of applying the quality control check of L&H equations (20) and (21), using the same assumed forecast error variance for all the observations. Figure 5(b) shows the results if the forecast error distribution is predicted by a perfect ensemble (i.e. the ensemble members were created by random sampling from a Gaussian with the correct variance for that observation's position). The lines with o symbols were obtained by putting the ensemble variance into the same L&H quality control equations. Even with an ensemble as small as 40 members, the use of an ensemble estimate of the variance for each observation leads to better decisions than just using the average variance[4]. As discussed in the previous section, an ensemble can represent a non-Gaussian forecast pdf. A Bayesian quality control algorithm can use the ensemble directly, replacing the integral over the pdf by a sum over ensemble members. The observation is rejected if

$$\sum_{i=1}^{N} P\left(y^o \middle| x_i \cap \overline{G}\right) P\left(\overline{G}\right) < \sum_{i=1}^{N} P\left(y^o \middle| x_i \cap G\right) P(G) . \qquad \textbf{(13)}$$

The left-hand side is evaluated assuming the observation does not have a gross error ($\overline{G}$), using the Gaussian observational error distribution, and the right-hand side just sums the constant pdf (0.043 hPa^{-1}), multiplied by the prior probability of a gross error (0.03). The results of this method are show by the lines with x symbols. As expected, many more members are needed to adequately define the pdf if no assumptions are made about its shape; in this case at least 400 members are needed for good results.

Checks such as this could easily be incorporated into a sequential EnKF; deciding whether to accept or reject each observation just before using it. But in this example only about a sixth of the erroneous observations could be detected - even a good knowledge of its variance cannot compensate for the forecast not being very accurate. Using nearby observations in a "buddy-check" is more reliable (L&H). Unfortunately such methods require all the observations to be processed together, while in a sequential EnKF algorithm early observations influence the quality control of later observations, but not

[4] This work only considers the benefits of ensemble predictions of covariances. It is possible that an ensemble would also provide a more accurate prediction of the mean, but this has not been demonstrated for short forecasts. So I am assuming the forecast value used in the conventional quality control of figure 5(a) equals the population mean in the ensemble methods of figure 5(b).

vice versa. So if the ensemble covariances are used in a quality control step within a sequential EnKF analysis algorithm, results will be sub-optimal.[5]

(f) Nonlinear observation operators

Nonlinear relationships between the observed quantities and the model variables are allowed for in the EnKF by (2) and (3); for quantities such as rainfall which are diagnosed during the model forecast they are allowed for by augmenting the $\{\mathbf{x}_i^f\}$ to include a rainfall rate variable. This makes it easy to design an EnKF which uses such observations, without understanding and linearising the processes involved, as would be necessary for a variational method.

But, even if the forecast error pdf is Gaussian, and can be described by $\mathbf{P}^f$, the action of nonlinear H gives a non-Gaussian distribution. The use of (2) and (3) does not allow for this, while variational methods with a non-quadratic penalty function can. The simplest example is an observation of a wind speed. Figure 6(a) shows an approximately Gaussian forecast ensemble of wind vectors, and figure 6(b) the resulting analysis ensemble. The joint pdf of (u,v) and windspeed is non-Gaussian, and therefore not fully described by (2), making the analysis using (6) sub-optimal. Hence the analysed ensemble does not represent the ideal posterior pdf, which should lie on the circle. In the subsequent analysis of other observations, this ensemble will be represented by the Gaussian whose mean and eigenvectors are the grey dot and cross. This sub-optimality probably does not matter for a single wind speed observation, particularly when realistic observation errors are included. But it might matter for some dense highly nonlinear observations such as scatterometer winds, where the detailed information in nearby observations can combine synergistically. Variational methods, allowing for nonlinear observation operators, accurately represent the pdf in the vicinity of the latest best estimate; in our example they would always give an analysed wind on the circle. They have proved useful (Stoffelen and Anderson 1997, Isaksen and Stoffelen 2000), allowing the analysis to implicitly select between wind aliases to give a horizontally consistent pattern.

5. Variational methods using ensemble covariances

(a) Variational fit to ensemble perturbations

It is possible to use the ensemble covariance (8), and its modified version using a Schur product, in the background term of a variational analysis scheme (either 3D-Var or 4D-Var). As discussed in section 4(b), $\mathbf{P}_e^f$ is degenerate, so we cannot immediately use it in a standard variational penalty function for $\mathbf{x}$. The null space of $\mathbf{P}_e^f$ constrains $\mathbf{x}$ to stay in the space spanned by the ensemble. Lorenc (2002) discussed how to incorporate such constraints by transforming a simpler control variable. In this case I choose the transformation from a vector $\boldsymbol{\alpha}$ of N coefficients α_i:

[5] If the ensemble is large enough to properly represent the tails of a non-Gaussian pdf, and if the nonlinear resampling methods mentioned in 4(d) are used, it is theoretically possible to construct an optimal sequential algorithm.

$$\mathbf{x} = \overline{\mathbf{x}^f} + \mathbf{X}^f \boldsymbol{\alpha} \,.$$

(14)

Then, if the covariance $\mathbf{B}_{(\alpha)} = \langle \boldsymbol{\alpha}\boldsymbol{\alpha}^T \rangle = \mathbf{I}$, the covariance of the $\mathbf{x}$ defined by (14) satisfies (8). So the variational analysis for the transformed variable $\boldsymbol{\alpha}$ is obtained by minimising the squared control variables plus the misfit between the predicted values $\mathbf{y}$ and observations $\mathbf{y}^o$:

$$J(\boldsymbol{\alpha}) = \tfrac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{\alpha} + \tfrac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right)$$
$$\mathbf{y} = H\left(\overline{\mathbf{x}^f} + \mathbf{X}^f\boldsymbol{\alpha}\right)$$

(15)

(b) Using the Schur product to reduce degeneracy

In order to match the covariance modified by the Schur product, we use as control variable a matrix $\boldsymbol{\alpha}$ whose columns are N fields $\boldsymbol{\alpha}_i$, each with the desired modifying covariance $\mathbf{C}$. The variational problem is then to minimise

$$J(\boldsymbol{\alpha}) = \tfrac{1}{2}\boldsymbol{\alpha}^T \begin{pmatrix} \mathbf{C} & & 0 \\ & \ddots & \\ 0 & & \mathbf{C} \end{pmatrix}^{-1} \boldsymbol{\alpha} + \tfrac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right)$$
$$\mathbf{y} = H\left(\overline{\mathbf{x}^f} + \mathbf{X}^f \circ \boldsymbol{\alpha}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\right)$$

(16)

where we use a similar Schur product[6] in calculating $\mathbf{x}$ and then $\mathbf{y}$ (see appendix B). As in standard 3D-Var methods (e.g. Lorenc *et al.* 2000), the inversion of the block diagonal matrix is avoided by a further horizontal transform of each field $\boldsymbol{\alpha}_i$ into spectral space. Note that where (14) multiplies each ensemble member by a scalar coefficient, (16) has a Schur product with a horizontal field of coefficients. This increases the degrees of freedom available to the variational algorithm. The fields $\boldsymbol{\alpha}_i$ are constrained to be smooth, with the horizontal scale in $\mathbf{C}$ 2-3 times that of forecast error covariances (figure 3). In response to a single observation, the fields $\boldsymbol{\alpha}_i$ will be proportional to the correlation function, dropping to zero away from the observation.

---

[6] In implementations to date $\mathbf{C}$ is constructed using a single horizontal correlation function, defining the correlation between any variables at any levels to be a function of position only. In a similar way the $\boldsymbol{\alpha}$ fields used in the Schur product in (16) are built up by replicating a single two-dimensional field of coefficients. In the control variable and first penalty term we need only consider the single horizontal field.

(c) Hybrid variational plus ensemble covariances

We could in principle minimise an equation like (16) N times, with perturbed observations, in the equivalent of (6), but this would be less efficient than the traditional EnKF method in generating an ensemble, since the saving obtained through the common term in (6) is not readily available. The main advantage is that we can have a hybrid approach, with the traditional variation control variable $\mathbf{v}$ (Lorenc 2002) supplemented by $\boldsymbol{\alpha}$. The incremental variational method can be used, where the control variables and the ensemble are at reduced resolution and $\mathbf{y}$ is given by equation (35) of Lorenc (2002). In a variational algorithm, whatever sequence of operations we use for calculating $\mathbf{y}$, we need its linearisation and adjoint to do the minimisation of

$$J(\mathbf{v},\boldsymbol{\alpha}) = \tfrac{1}{2}\mathbf{v}^T\mathbf{v} + \tfrac{1}{2}\boldsymbol{\alpha}^T \begin{pmatrix} \mathbf{C} & & 0 \\ & \ddots & \\ 0 & & \mathbf{C} \end{pmatrix}^{-1} \boldsymbol{\alpha} + \tfrac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right), \qquad \textbf{(17)}$$

where, as noted above, a further horizontal transform of $\boldsymbol{\alpha}$ removed the need to invert $\mathbf{C}$. The ensemble perturbations can come from another source such as an error breeding cycle (Toth and Kalnay 1997). It has proved relatively straightforward to code this within the Met Office variational assimilation, with encouraging initial results (Dale Barker and Adrian Semple, personal communication). A similar hybrid method, not using the localising Schur product, has been demonstrated on a simple model by Hamill and Snyder (2000).

The hybrid method is easily extended from 3D-Var to 4D-Var by inserting a forecast step into the calculation of $\mathbf{y}$, as in normal 4D-Var schemes. Note that in this case the ensemble members are still used at a single time at the beginning of the window; the four-dimensional covariances discussed in section 2(c) are not used. Within the window the time-evolution of covariances is modelled by 4D-Var, and the EnKF models their evolution from one window to the next.

The variational methods in this section are able to use nonlinear observation operators, and variational quality control methods which simultaneously check all observations, while using covariances (partly) modelled by the ensemble, giving the advantages discussed in 4(e). without the sub-optimal sequential processing.

6. Discussion and Conclusions

In section 3 I compared the easy of implementation of the EnKF and 4D-Var (summarised in Table 1), concluding that, for some aspects, the design of an EnKF scheme is simpler. For ensemble sizes O(100) costs should be similar. So using an EnKF is attractive for a centre which does not have existing 4D-Var infrastructure. It is also attractive in generating an ensemble, which will be useful in providing probability forecasts to users.

But this attractiveness will be lost if better estimates of the forecast errors are at the expense of degradation to the best estimate analysis and forecast, so the characteristics of the EnKF discussed in section 4 are crucial. Table 2 summarise these and compares with those of 4D-Var.

Figure 2 shows that the deleterious effect of sampling noise on the covariances can be ameliorated by an appropriate Schur product, for an ensemble size of 100~1000. Current forecast models represent well the dynamical growth of mid-latitude weather systems over periods of a few days, and the EnKF should model the consequential error covariances, which depend on the synoptic flow, better than 4D-Var (which is limited to modelling the linear evolution within the 4D-Var window). So for the mid-latitude medium-range forecasting problem, where such weather systems are important, and where ensemble predictions have proven utility, the EnKF is attractive. Such a system is likely to use sequential processing of batches of observations, and an ensemble size O(100) (Houtekamer and Mitchell 2001). With this design, it will not be able, directly and optimally, to use observations with non-Gaussian errors or nonlinear observation operators. The method can use such data simply and directly in a sub-optimal way; alternatively a separate pre-processing step might use the ensemble covariances, for instance for quality control and scatterometer de-aliasing (section 4(d) (e) and (f)). The ensemble should usefully predict day-to-day variations in the (co)variances used in pre-processing these observations, but it is unlikely the ensemble will be large enough to accurately represent non-Gaussian properties of the forecast errors (figure 5(b)). A remaining worry for the application of the EnKF to medium-range forecasting is that, even for synoptic-scale forecasts, important information may be present in observations which require high-resolution assimilation, for instance the tracer observations discussed below. Improved assimilation of satellite data in a high-resolution 4D-Var system has been an important factor in recent improvements in medium-range forecasts (Simmons and Hollingsworth 2002); the use of the same data in an EnKF has yet to be demonstrated.

The limit a finite ensemble size puts on the ability to fit detailed observations may be a significant problem for some applications. The atmosphere has many superimposed scales of motion; even for mesoscale assimilation systems, synoptic scale errors occur, and the dynamical error covariance scales are several hundreds of kilometres. The scale of the Schur product correlation should be 2-3 times this to get accurate dynamical analyses, as shown in figure 3. But the scales which we wish to analyse in humidity fields are much smaller, and from satellite and radar we may have many hundreds of observations within the Schur correlation radius. Providing a good fit to these observations is important for at least two reasons:

- to have accurate high-resolution cloud and precipitation at the beginning of the a short-period mesoscale forecast. (If the "nowcast" does not correctly represent the current weather, customers lose confidence in the forecast!)
- to use the information from the movement seen in consecutive observations to deduce advecting velocities. This effect has been demonstrated in 4D-Var mesoscale assimilation of radar reflectivity, and in global 4D-Var assimilation of satellite soundings. The Kalman filter is similarly capable of extracting information about the advective velocity (Daley 1996), but only if the analysis fields can represent the pattern seen in the first set of observations. To extract this tracer information, the assimilation needs to process frequent observations. The EnKF may also have some practical disadvantages in this, because of the need for initialisation (discussed in 4(c)), and also the transposition costs (discussed in 3(c)).

Most NWP systems of the future will have to assimilate remote sensing observations with high spatial and temporal resolution to produce detailed analyses, while still using sophisticated methods to fill in incomplete information[7]. 4D-Var can do this. It is not clear that the EnKF, with its limited ability to fit detailed observations without using a very large ensemble, will suite this scenario as well as variational methods. It may be better to consider the hybrid methods described in section 5, which use 4D-Var to model the evolution within each window, and can thus properly extract tendency information from observation in the same window. In order to run high-resolution NWP systems, most centres will continue to use limited-area models[8]. As discussed in section 3(d), boundary conditions from a global ensemble are needed when generating a limited-area ensemble; it cannot simply perpetuate itself using the EnKF methods of section 2. So the hybrid methods look particularly attractive for limited-area, mesoscale assimilation.

Table 1 Summary of the comparative ease of implementation of the ensemble Kalman filter, and 4D-Var (discussed in section 5).

Forecast model	Predict evolution of mean. No switches.	Predict evolution of typical state. May have stochastic physics and switches.
Linear model	Predict average evolution of finite perturbations from the mean. May be simplified.	Not needed.
Adjoint model	Needed for (simplified) linear model.	Not needed.
Suitability for parallel computers	Require parallel simplified and adjoint models.	Ensemble members can run on individual processors, but then requires a costly data transposition for the covariances.
Simplicity of design and coding	Significant effort needed for covariance model, and for suitable linear approximations to nonlinear model. Adjoint code needed.	Little effort, given existing forecast model.
Limited-area modelling	Error covariance models can be extended to specify boundary value errors	Requires an ensemble of global forecasts to provide boundary conditions.

[7] We cannot expect to have sufficient observations to define directly what we need to know, since observations are relatively expensive compared to data assimilation systems, so it is more cost effective to reduce the observations and have a sophisticated assimilation.

[8] Others may avoid the boundary problem using a stretched global grid as in Côté *et al.* 1998.

Table 2 Summary of the assimilation characteristics of the ensemble Kalman filter (discussed in section 6), compared with those of 4D-Var (discussed in Lorenc 2002).

	Incremental 4D-Var	EnKF
Forecast covariances	Modelled at t_0 (usually isotropic), time evolution represented by linear and adjoint models.	Sampled by ensemble (flow-dependent). Noisy: must be modified to have compact support using Schur product.
Analysis method	Variational minimisation using model increments, with full model outer-loop.	Observation space sequential KF eqn, modified to cope with degenerate sample covariances.
Ability to fit detailed observations.	Limited by resolution of simplified model.	Limited to fewer data (in a region) than ensemble members.
Balance constraints	Can be imposed through a dynamical design to the variable transform, or a separate balance penalty.	Only imposed if each forecast in the ensemble is balanced. Lost slightly in Schur product.
Nonlinear observation operators	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Allowed, but resulting pdf modelled by Gaussian.
Non-Gaussian observational errors	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Not allowed. Prior QC step is needed.

Legends for figures

Figure 1. Ensemble estimate covariances with a point at the origin, for ensembles of 100 samples taken from a population whose covariance, shown in the solid curve, is $\exp\left(-r^2/2a^2\right)$, with $a = 500$ km.

Figure 2. The change in the global average error variance, calculated using (12), for ensemble covariances like those in figure 1, modified by a Schur product with a correlation **C**. The change is plotted as a function of the relative scale of the correlation function **C**, such that 1 implies that **C** is similar to the covariance curve shown in figure 1. A symbol marks the best scale for **C** (by this measure), different curves are for different ensemble sizes, and the dashed line shows the error reduction achievable with perfect covariances.

Figure 3. As figure 1, with the ensemble covariances modified with the Schur product correlation **C**, with scale corresponding to the symbol in figure 2 (shown dashed). The Schur product of **C** with the population covariance is shown in the lower grey line.

Figure 4. Factor by which, because of the Schur product, the analysed wind will be less than that required to balance the analysed height gradient, at 500 km from a single height observations, as a function of ensemble size. The best scales for **C** shown in figure 2 were used.

Figure 5. Percentage of all observations correctly identified as having gross errors (solid lines), and falsely identified (dotted lines). (a) using the Bayesian QC algorithm of Lorenc and Hammon (1998), for a range of assumed background error variances. (b) lines with o use the variance from an ensemble in the same algorithm, and lines with x use the ensemble pdf directly with no assumption about its shape, for a range of ensemble sizes.

Figure 6. (a) 1000 member ensemble of forecast wind vectors, with mean (2,4) and standard deviation (2,2). (b) Ensemble of wind vectors resulting from an EnKF analysis of an error-free observation of wind speed; the circle shows the observed value, 3. The large grey dots show the mean of each distribution, and the grey lines in (b) show its eigenvectors.

Appendix A

By subtracting (4) and (6), and then substituting (8), it is easy to show that

$$\mathbf{x}_i^a - \overline{\mathbf{x}^a} = \sum_{j=1}^{N} \left(\mathbf{x}_j^f - \overline{\mathbf{x}^f} \right) w_{j,i} \quad for\ i = 1, N\,. \tag{18}$$

If the analysis has just assimilated an error free observation, all the analysed ensemble members, and the mean, will fit the observation. Applying the (linear) observation operator H to (18) gives

$$0 = \sum_{j=1}^{N} \left(H\mathbf{x}_j^f - H\overline{\mathbf{x}^f} \right) w_{j,i} \quad for\ i = 1, N\,. \tag{19}$$

So, as long as the $H\mathbf{x}_j^f$ are different, the matrix of $w_{j,i}$ is degenerate, and adding the error-free observation will remove a degree of freedom from the ensemble.

Appendix B

Consider element (l,m) of the covariance of $\mathbf{x}$ given by

$$\mathbf{x} = \overline{\mathbf{x}^f} + \mathbf{X}^f \circ \boldsymbol{\alpha} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{20}$$

We show that this is the same as element (l,m) of $\mathbf{P}_e^f \circ \mathbf{C}$. I use x_i' to denote the i'th column of $\mathbf{X}^f$:

$$\begin{aligned}
\left\langle \left(x(l) - \overline{x^f(l)}\right)\left(x(m) - \overline{x^f(m)}\right)\right\rangle &= \left\langle \sum_{i=1}^{N} \alpha_i(l) x_i'(l) \sum_{j=1}^{N} \alpha_j(m) x_j'(m) \right\rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i'(l) x_j'(m) \left\langle \alpha_i(l) \alpha_j(m) \right\rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i'(l) x_j'(m) C(l,m) \delta_{ij} \\
&= \sum_{i=1}^{N} x_i'(l) x_i'(m) C(l,m) \\
&= P_e^f(l,m) C(l,m) \\
&= \left(\mathbf{P}_e^f \circ \mathbf{C}\right)_{l,m}
\end{aligned} \tag{21}$$

References

Anderson, Jeffrey L. 2001: "An ensemble adjustment Kalman Filter for data assimilation" *Mon. Wea. Rev.*, **129**, 2884-2903

Andersson, Erik and Jarvinen, Heikki. 1999: "Variational Quality Control" *Quart. J. Roy. Met. Soc.*, **124**, 697-722

Andersson, E., J. Pailleux, J-N. Thepaut, J. R. Eyre, A. P. McNally, G. A. Kelly, and P. Courtier 1994: "Use of cloud-cleared radiances in three/four-dimensional variational data assimilation" *Quart. J. Roy. Met. Soc.*, **120**, 627-653

Bishop, C. H., B. J. Etherton and S. J. Majumdar 2001: "Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects", *Mon. Wea. Rev.*, **129**, 420-436.

Burgers, G., P. J. van Leeuwen and G Evensen, 1998: "Analysis scheme in the ensemble Kalman filter." *Mon. Wea. Rev.*, **126**, 1719-1724

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: "The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation". *Mon. Wea. Rev.*, **126**, 1373-1395.

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A Hollingsworth, F. Rabier and M. Fisher. 1998: "The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation" *Quart. J. Roy. Met. Soc.*, **124**, 1783-1808

Daley R. 1996: "Recovery of the one and two dimensional windfields from chemical constituent observations using the constituent transport equation and an extended Kalman filter." *Met. and Atmos. Phys.*, **60**, 119-136.

Dharssi, I., Lorenc, A. C. and Ingleby, N. B. 1992: "Treatment of gross errors using maximum probability theory" *Quart. J. Roy. Met. Soc.*, **118**, 1017-1036

Evensen, G., 1994: "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics." *J. Geophys., Res.*, **99**(C5), 10143-10162.

Gaspari, G. and Cohn, S. E. 1999: "Construction of correlation functions in two and three dimensions." *Quart. J. Roy. Met. Soc.*, **125**, 723-758

Gauthier, P. and J.-N. Thépaut, 2001: Impact of the Digital Filter as a Weak Constraint in the Pre-operational 4DVAR Assimilation System of Météo-France. *Mon. Wea. Rev.*, **129**, 2089-2102.

Hamill, T. M. and C. Snyder, 2000: "A hybrid ensemble Kalman filter - 3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

Hamill, T. M., J. S. Whitaker and C. Snyder, 2001: "Distance dependent filtering of background error covariance estimates in an ensemble Kalman filter." *Mon. Wea. Rev.*, 129, 2776-2790.

Houtekamer, P. L. and H. L. Mitchell, 2001: "A sequential ensemble Kalman filter for atmospheric data assimilation." *Mon. Wea. Rev.*, **129**, 123-137

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C. 1997: "Unified notation for data assimilation: Operational, Sequential and Variational" *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 181--189

Isaksen, L. and A. Stoffelen, 2000: "ERS-scatterometer wind data impact on ECMWF's tropical cyclone forecasts" ECMWF Tech Memo 316.

Keppenne, C. L. 2000: Data Assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981

Lorenc, A.C. and Hammon, O., 1988: "Objective quality control of observations using Bayesian methods. Theory, and a practical implementation." *Quart. J. Roy. Met. Soc.*, **114**, 515-543

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne and F. W. Saunders. 2000: The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Quart. J. Roy. Met. Soc.*, **126**, 2991-3012.

Lorenc, A. C. 2002: Modelling of error covariances by four-dimensional variational data assimilation. *In preparation.*

Parrish, David F., and Derber, John C. 1992: "The National Meteorological Center's Spectral Statistical-Interpolation analysis system" *Mon.Wea.Rev.* **120,** 1747-1763

Rabier, F., H. Jarvinen, E. Klinker, J-F. Mahfouf, A. Simmons. 2000: The ECMWF operational implementation of four-dimensional variational assimilation - Part I: experimental results with simplified physics. *Quart. J. Roy. Met. Soc.*, **126**, 1143-1170.

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **128**, 647-677

Stoffelen, A. and D. Anderson, 1997: "Ambiguity removal and assimilation of scatterometer data", *Quart. J. Roy. Met. Soc.*, **123**, 491-518

Toth, Z. and E. Kalnay, 1997: "Ensemble forecasting at NCEP and the breeding method". *Mon. Wea. Rev.*, **125**, 3297-3319.

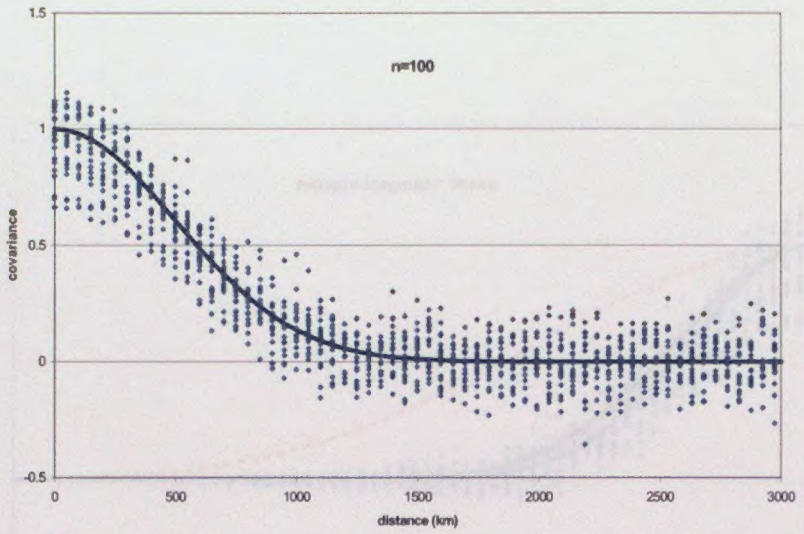

Figure 1. Ensemble estimate covariances with a point at the origin, for ensembles of 100 samples taken from a population whose covariance, shown in the solid curve, is $\exp\left(-r^2/2a^2\right)$, with $a = 500$ km.

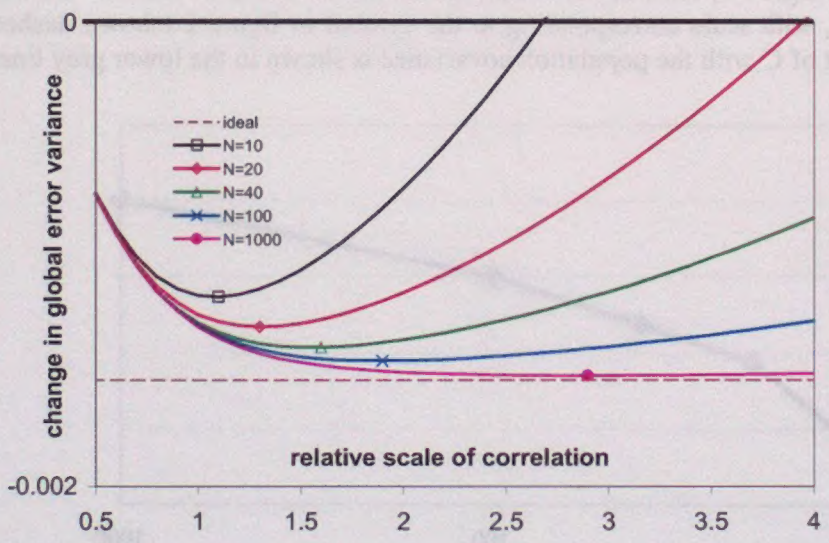

Figure 2. The change in the global average error variance, calculated using (12), for ensemble covariances like those in figure 1, modified by a Schur product with a correlation **C**. The change is plotted as a function of the relative scale of the correlation function **C**, such that 1 implies that **C** is similar to the covariance curve shown in figure 1. A symbol marks the best scale for **C** (by this measure), different curves are for different ensemble sizes, and the dashed line shows the error reduction achievable with perfect covariances.

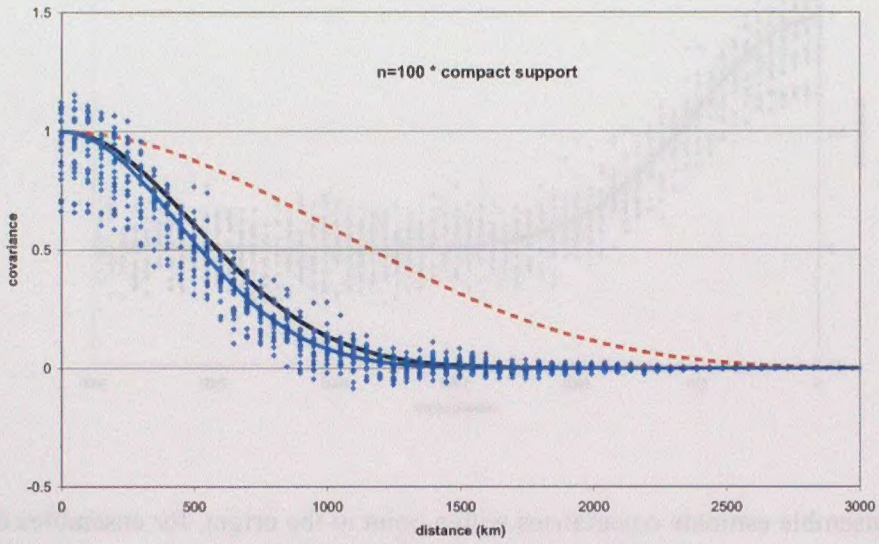

Figure 3. As figure 1, with the ensemble covariances modified with the Schur product correlation **C**, with scale corresponding to the symbol in figure 2 (shown dashed). The Schur product of **C** with the population covariance is shown in the lower grey line.

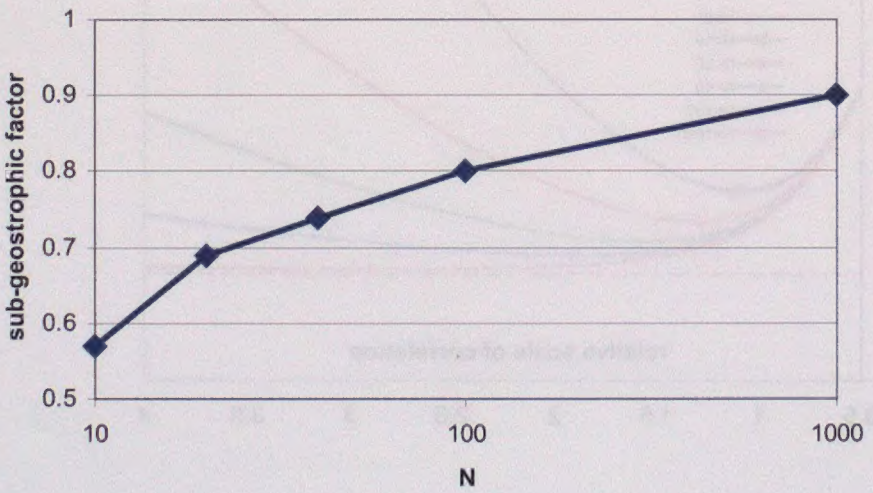

Figure 4. Factor by which, because of the Schur product, the analysed wind will be less than that required to balance the analysed height gradient, at 500 km from a single height observations, as a function of ensemble size. The best scales for **C** shown in figure 2 were used.

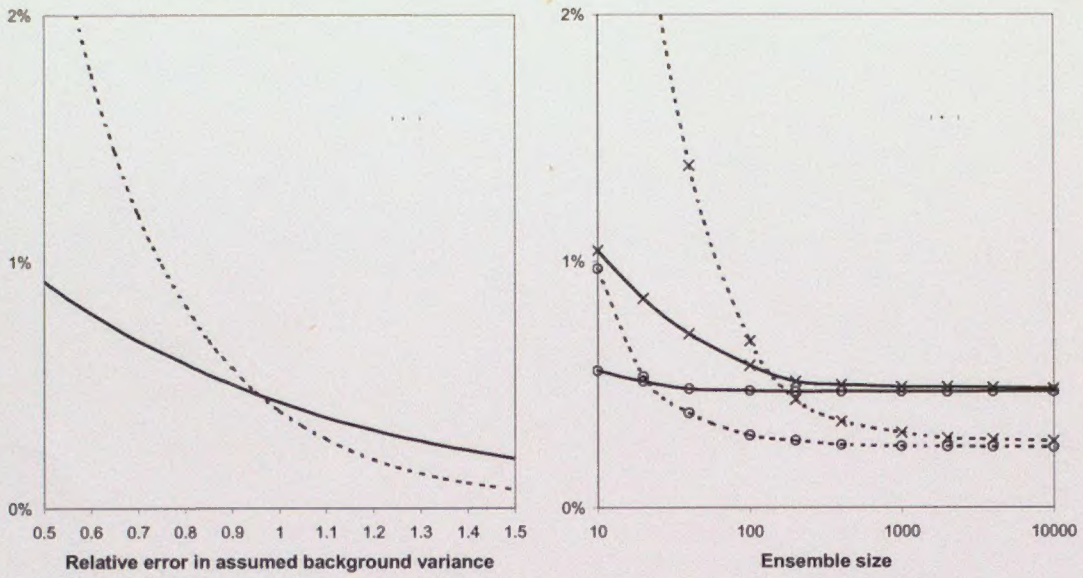

Figure 5. Percentage of all observations correctly identified as having gross errors (solid lines), and falsely identified (dotted lines). (a) using the Bayesian QC algorithm of Lorenc and Hammon (1998), for a range of assumed background error variances. (b) lines with o use the variance from an ensemble in the same algorithm, and lines with x use the ensemble pdf directly with no assumption about its shape, for a range of ensemble sizes.

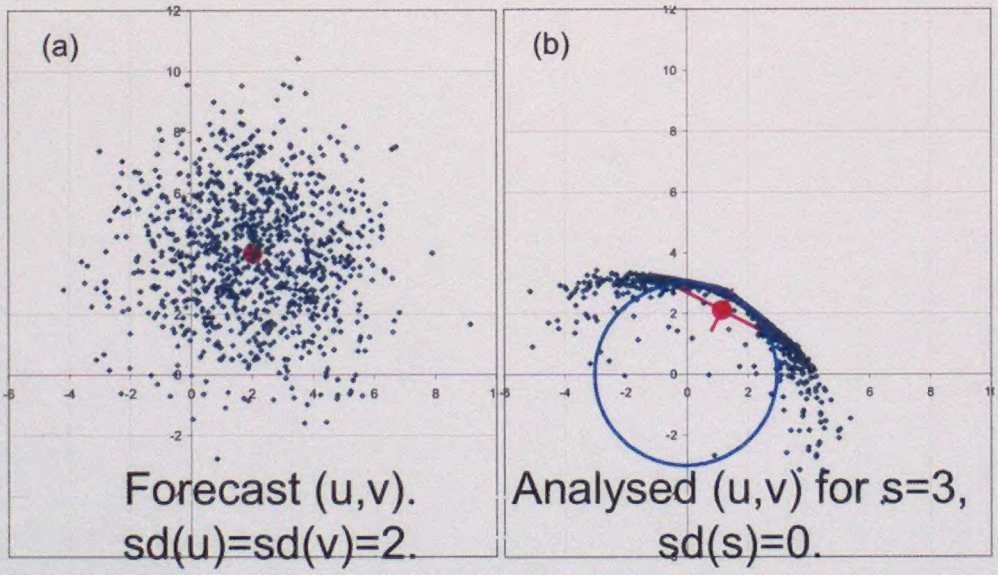

Figure 6. (a) 1000 member ensemble of forecast wind vectors, with mean (2,4) and standard deviation (2,2). (b) Ensemble of wind vectors resulting from an EnKF analysis of an error-free observation of wind speed; the circle shows the observed value, 3. The large grey dots show the mean of each distribution, and the grey lines in (b) show its eigenvectors.